

CMB91-547
& CMU

高等医学教育现代考试方法

——美国中华医学基金会认知能力评价项目的成果

主编 陈永祥 金魁利
著者 Leon J. Norcini, Ph.D.

主 编 孙宝志 金魁和

顾 问 John J. Norcini, Ph. D.

编 者 弓嫣嫣 王 威 姚 江 刘羽时
张训巍 蒋本铁 乔 敏 于晓松
时 谨 张志东 王 烈 邹 勇
马述茵 应宝俊 姚 颖 喻春英
郝永强 左天明

序

——美国中华医学基金会主席 威廉·索耶博士

近几十年来，一些国家的教育工作者已经相当重视对学生的知识和推理过程进行客观地定量地评价。一个实用的教育目标分类学的兴起指导了考试的建构。完美的统计学分析对试题的有效性和考试的有效性进行检验，以及为考试决策服务。计算机技术的应用，有力地促进了考试评分、分析和大型有效的分门别类的题库的开发等等，这些已经被考试管理者应用于考试的准备，有时也被学生用于自我评价。

我与许多有经验的教育工作者的共识是（1）课程能指导教师的教学活动；（2）评价能指导学生的学习活动。这是对复杂的教与学过程的最简单的表述。我们大家都熟悉学生常常问的一个问题，“这个将来考吗？”，的确对他们来说，评价是对学生的学习行为的强有力的向导。它是关键，因此评价的手段应当最合理。现代评价技术推动我们去实现这个目标。

在许多国家，不同种类的专业执照和资格证书的颁发主要依靠考试。因此社会和应试者都有权希望考试能质量好，公正合理，而且涉及有关问题和及格线确定都适当标准化。现代评价技术帮助这些希望成为现实。

本书叙述了作者们在中国运用现代考试技术与方法的实践。他们论述的目标、问题和成就对其他寻求改善对学生的知识和推理过程评价的人们具有相当重要指导的作用。

（签名）William D. Sawyer, M.D.

1996年8月7日于美国纽约 （孙宝志译）

FOREWORD

During recent decades, educators in several nations have directed considerable attention to the objective, quantitative assessment of students' knowledge and reasoning. A useful taxonomy emerged and guided the construction of examinations. Elegant statistical analyses provided tests of the validity of test items and examinations and served to guide decisions such as pass and fail. Computer technology greatly facilitated both the scoring and analysis of examinations and the development of large banks of validated and categorized test items. The latter have been used by those responsible for preparing examinations and sometimes, by students for self-assessment.

I share the belief of numerous experienced educators that (i) the curriculum serves primarily as a guide for faculty activities, and (ii) the evaluations are the primary guide to student learning. While this may be an oversimplification of the complex processes of education and learning, all of us are familiar with students' common question, "Will this be on the test?" Surely to them, the evaluation is a strong clue that directs their study and learning behavior. It is critical, therefore, that evaluation instruments be optimal. Modern evaluation technologies propel us to that goal.

In many nations, diverse kinds of licenses and certifications depend primarily on examinations. Society as well as the examinees legitimately expect the examinations to be of quality, fair, and suitably standardized with regard to relevance and pass-fail

decisions. Modern evaluation technologies help to make that expectation a reality.

This book describes the efforts of the authors to employ modern evaluation strategies and technologies in China. Their description of objectives, difficulties, and achievements can guide others who seek to improve their evaluations of their students' knowledge and reasoning.

William D. Sawyer, M. D.
President
China Medical Board
of New York, Inc.

August, 1996

前 言

几乎所有的教学过程,包括高等医学教学过程,都存在考试这一重要环节。没有考试,教学过程就意味着不完整。考试的主要目的是在重要的行为领域内获得信息,以便对受试者的素质和能力作出判断。

在教学过程中,教学开始阶段,安排摸底测验和配置性测验,以掌握学生的知识与能力的水平,做到心中有数,教学有的放矢。教学中间阶段安排形成性测验和诊断性测验,以便帮助学生,诊断学习中存在问题。教学结束后,安排总结性测验或称总结性考试,鉴定学生个体的学习效果,以便升级和毕业注册成绩;反馈教学效果,以便教师改进教学工作;为学校控制教学质量提供信息依据等。国外,高等医学教育考试还要为医师执照的颁发提供合格的证据。因此,本书的书名中现代考试方法的前面冠以高等医学教育几个字,就是强调考试在教学过程中的地位与作用。本书名中现代考试方法几个字中“现代”的含义是对过去的传统方法比较而言。

现代科学技术飞跃发展,也促进了考试方法的改进,从19世纪末到20世纪末以来,传统经验考试发展到现代的科学化考试。

在考试内容方面,由考核记忆知识变为侧重考核理解和运用,注重理论紧密联系实际。

在考试方法方面,既继承和改造传统方法,又创造与现代考试内容和技术手段相适应的新方法。在笔试方面产生了能对测试误差进行有效控制的标准化考试,并使用电子计算机阅卷判分,统计分析,提高了考试的效度和信度,这是现代考试的重要标志。为适应现代科技需要,实践考试方法也更接近实际,出现模拟病人

考试、模拟计算机考试,模型病人计算机程序考试等,把现代化技术应用于考试中。

在考试管理方面,运用现代信息科学技术、组织大规模的考试,更加科学,更加严密。

这些现代考试的特点,在美国专家的指导下,我们经过5年的实践,在本书中有所体现。

关于本书副标题“实施美国中华医学基金会认知能力评价项目的成果”,在本书第一章绪论中将介绍项目的立项依据,中美比较与改革的目标,项目过程与成果,以便读者对本书的理论体系和经验给予理解。

本书第二章中国与美国高等医学教育考试的比较与改革目标,第三章介绍教育目标分类学理论,以便为现代考试设计,提供理论依据。第四、五、六章,介绍笔试试题的编制新技术,第七章介绍现代临床实践考试方法,第八章介绍计算机模拟考试,这些都体现现代医学考试方法,也反映出世界前沿技术。第九、十、十一章,介绍考试管理,医学考试分析、医学试题分析的方法,第十二、十三章介绍医学考试及格标准确定技术,医学考试等值化技术,这些都是国内以前没有应用过的最新技术。第十四章介绍医学考试研究,这是为医学考试服务的重要措施,是体现现代化考试的特点。第十五、十六章介绍临床医学综合考试国家题库建设过程与成果,很多方面属国内首创,达到国际水平并通过国家级科技成果鉴定。题库已由高等教育出版社正式发行,推广到全国70%的医学院校。第十七章介绍国内外的医师考试试题范例,以供教师和学生参考。

本书由孙宝志、金魁和担任主编,美国国家内科委员会副主席 John. J. Norcini, Ph. D. 担任本书的顾问,美国中华医学基金会主席 William. D. Sawyer, M. D. 给本书以指导帮助。全国17所医学院校在项目过程中1994年至1995年参加临床医学专业联合考试,对项目工作给予大力支持,在此表示衷心感谢。这些学校是:

大连医科大学、上海第二医科大学、上海医科大学、山东医科大学、中国协和医科大学、中国医科大学、石河子医学院、白求恩医科大学、北京医科大学、福建医学院、华西医科大学、首都医科大学、重庆医科大学、第一军医大学、浙江医科大学、湖北医科大学、湖南医科大学。

本书各部分编写人员是：第一章，金魁和；第二章，孙宝志；第三章，孙宝志，王威；第四章，于晓松，孙宝志；第五章，孙宝志，于晓松，邹勇；第六章，应宝俊，喻春英，王威；第七章，孙宝志，乔敏；第八章，刘羽时，乔敏；第九章，姚江；第十章，弓嫣嫣，时谨；第十一章，弓嫣嫣，张训巍；第十二章，弓嫣嫣，姚颖；第十三章，张志东，王烈，马述茵；第十四章，孙宝志，弓嫣嫣，郝永强；第十五章，孙宝志；第十六章，蒋本铁，周福才；第十七章，张训巍，左天明等。

本书编写过程中，很多章节，都是培训讲义改写的，经过反复修改。有些材料是引进国外技术编写的。由于水平有限，编写仓促，错误之处在所难免，敬请读者批评指正。

孙宝志 金魁和

于沈阳 中国医科大学

1996年8月

目 录

序 美国中华医学基金会主席威廉·索耶博士	
前言	1
第一章 绪 论	1
第一节 我国高等医学教育考试改革(认知能力评价项目)的立项依据	2
第二节 我国高等医学教育考试(认知能力评价项目)的过程与成果	4
第二章 中国与美国高等医学教育考试比较与改革目标	8
第一节 美国高等医学教育考试体系	8
第二节 中国高等医学院校考试现状调查	14
第三节 中国高等医学教育考试改革要解决的主要问题	22
第三章 教育目标分类学在医学教育考试中应用	27
第一节 教育目标分类学简介	28
第二节 医学教育中的教育目标分类理论	32
第四章 标准化笔试——医学多选题新类型	44
第一节 最佳选择题	44
第二节 配伍题	49

第五章 医学多选题命题新技术	51
第一节 医学多选题命题原则	51
第二节 医学多选题命题新技术	60
第三节 临床实践型多选题编写技术	69
第六章 传统笔试——医学自由应答试题新技术	77
第一节 简答题	77
第二节 论述题	81
第三节 填充题	86
第四节 判断说明题(改错题)	88
第五节 改进型论述题	90
第七章 现代临床实践考试方法	94
第一节 临床实践考试的基本问题	94
第二节 标准化病人考试	102
第三节 客观结构化临床考试	110
第八章 计算机模拟考试	120
第一节 计算机模拟考试的概况	121
第二节 计算机模拟病例程序考试	125
第三节 计算机模拟考试的评分方法	135
第四节 计算机模拟考试的优缺点	137
第九章 医学考试管理	139
第一节 我国医学考试管理的现状	139
第二节 医学考试组织机构的人员与素质	141
第三节 医学考试计划管理	143
第四节 医学考试实施管理	159

第十章 医学考试分析	182
第一节 考试分数的解释.....	183
第二节 测量误差.....	187
第三节 考试效度.....	189
第四节 考试信度.....	196
第五节 考试分析对改进教学的反馈.....	207
第十一章 医学试题分析	209
第一节 试题内容评价.....	209
第二节 试题的难度.....	212
第三节 试题的区分度.....	223
第四节 试题参数.....	229
第五节 试题偏差.....	232
第十二章 确定考试及格标准新技术	239
第一节 判断试题内容制定及格标准.....	240
第二节 判断考生个体制定及格标准.....	246
第三节 用安国辅方法制定医学考试及格线的实验研究与应用.....	250
第十三章 医学考试等值化技术	265
第一节 等值化的概念、原理和方法	266
第二节 等值化方法注意事项.....	275
第三节 题库建设中试题难度等值化.....	278
第十四章 医学考试研究	291
第一节 医学考试研究的原则与方法.....	291
第二节 医学考试研究范例.....	292

第十五章 临床医学综合考试国家题库——总体设计与成果	316
第一节 计算机化题库建设国内外进展	316
第二节 医学考试题库设计思想	319
第三节 医学综合考试题库研制技术路线与成果	323
第四节 临床医学专业综合考试国家题库全貌	335
第十六章 临床医学综合考试国家题库——计算机软件系统	338
第一节 系统结构与功能	338
第二节 成卷原理及设计思想	340
第三节 题库系统的安装与操作方法	347
第十七章 国内外医师考试试题范例	376
第一节 介绍美国国家医师执照考试大纲目录及例题	376
第二节 我国临床医学专业综合考试国家题库考试大纲及部分试题	492

第一章 绪论

当今世界国际竞争日趋激烈,科学技术发展迅速。世界范围的经济竞争、综合国力竞争,实质上是科学技术的竞争和民族素质的竞争。从这个意义上说,谁掌握了面向二十一世纪的教育,谁就能在二十一世纪的国际竞争中处于战略主动地位。

世界范围的新技术革命的兴起,导致产业结构、技术结构、职业结构和人才素质结构的深刻变革。迅速提高劳动者的素质和智能,准确测验劳动者的质量,恰当选任符合现代化的人才,日益成为决定社会生产发展进程的关键。因此,本世纪70年代末期,一些发达国家便把考试看作迎接新技术革命挑战战略研究中重大理论课题。现阶段、考试学已是国外探索的一门新兴科学,其理论体系正在形成之中。医学考试学作为考试学的分支,也在积极酝酿之中,还没有形成完整的学科体系。

中国医科大学于1991年11月获得美国中华医学基金会资助大型项目CMB91-547《认知能力评价》,该项目从1991年11月起至1996年11月止,共计5年时间。引进国外的先进经验,研究具有中国特色的医学考试认知领域的理论和方法体系,与国际标准接轨。对我国高等医学教育,改革与发展有重大意义。

本项目在美国中华医学基金会主席索耶博士(William D. Sawyer, M. D.)和项目顾问诺齐尼博士(John J. Norcini, Ph. D.)的指导下,经过以中国医学大学医学教育研究中心领导的课题组专家的共同努力下,顺利地完成了5年工作任务,取得良好效益,初步形成了一套在中国可行的医学考试认知领域理论与方法;建立了计算机题库智能组卷系统,自动阅卷应用系统,考试分析系统等。向考试命题科学化、考试管理系统化、考试分析自动化、

考试研究理论化方向,迈出了成功的一步。

本书就是该项目五年研究与实践的理论结晶。为了使读者更好地阅读此书,在本章绪论中首先介绍一下《认知能力评价》项目的立项依据及整个项目活动过程。使大家了解项目的目标、方法、过程和结果,更有利于对我国高等医学教育考试改革的认识,也能更好地运用高等医学教育现代化考试方法,以保证为社会培养高素质医学人才的质量。

第一节 我国高等医学教育考试改革(认知能力评价项目)的立项依据

中国的高等医学考试体制有自己的特色与优势,几十年来为中国培养了大批的合格医生,但是按四个现代化要求,与发达国家,特别是与美国相比,差距还是很大的。(这里中国方面专指培养医生体制,不包括研究生体制)

首先从体制上看:

1. 美国医学考试体制有三大块。一块由医学院校负责,一块由国家医学执照考试委员会负责,一块由国家医学专业考试委员会负责。这三大块构成连续统一体。

中国的医学考试体制(从培养医生角度)仅有一块,就是由医学院校自行负责,不管学校大小、质量高低、一律自行考试。很难控制医生教育质量。

2. 美国学位考试与医师执照考试分离。学位考试由院校负责,医师执照考试由国家医学考试委员会负责,由州政府颁发行医执照。

某种意义说,中国医学学位考试与医师执照考试二者合一,学生从医学院毕业就取得学位证书和毕业证书,也就相当于医生执照,可以在中国独立行医。严格地说,中国没有明确的医师执照考

试制度。

3. 美国医学教育年限长,入学标准大学毕业起点,医学院学习4年,毕业后住院医训练3—7年。

中国高中毕业直接入医学院,医学院学习5年(本科),毕业后住院医教育长短不一。

4. 美国医学院招生,举行全国统一医学院入学考试。

中国医学院招生随全国其它各类院校一起招生考试,没有自己的特点。

5. 美国有各专业医师考试,即内科医师有内科医师证书考试、外科有外科医师证书考试,这和行医执照两码事,行医执照是参加考试的必要条件。

中国没有专业医师证书考试。

6. 美国专业医师有再鉴定证书考试,如内科医师十年以后需再鉴定考试,主考医学新进展内容。

从方法上看:

在美国,考试与教育相对独立,相互制约;无论是医学院校,还是国家考试委员会,都建立各级考试委员会,(专家命题委员会,审题委员会);进行考试科学研究与应用;运用现代教育测量理论,指导医学考试;制定详细的考试计划、考试大纲;建立考试题库;电子计算机管理考试、阅卷、统计分析等;考试结果反馈给应试者,也反馈给教学系统。

中国现状如何?

中国医大医学教育研究中心,曾于1992年对全国现有的132所医学院校进行了摸底调查。有109所学校回答问卷,占全国83%,包括21所中医院校,15所医学专科,73所本科医学院校。调查结果看,中国在专家命题,建立卡片题库,考试分析等有一定基础,但是差距也相当大,主要表现在以下几点:

1. 按教育目标分类学指导命题的院校仅占10%;

2. 建立各学科考试命题大纲的院校仅占26.6%;一些院校学

科考试命题完全由教研组负责,教研组下放权力给授课教师,教师就按教学“讲什么,考什么,”很难控制考试质量。

3. 对考试进行常规统计分析的院校占 50%,对考试有效性分析的占 33%,对可靠性分析的占 37%,有难度指数分析的占 57%,有试题区别度分析占 43%。

4. 建立试题库,包括卡片式题库与计算机题库占 30%。

5. 运用电子计算机管理考试(如阅卷、题库、考试分析)的院校仅占 22%。

6. 有些学校对考试进行研究和反馈教学不重视,往往注重考试分数存入档案,升级,毕业,忽视用于改进教学。

7. 某些院校考试分数贬值,绝对分数偏高。原因是近年来,某些政府人事部门动机是好的,为了杜绝走后门,按学生毕业考试成绩高低,分配高低工作,但是客观上形成不同学校考试分数高低影响毕业分配去向,造成不管学校水平如何,只注重绝对分数,这完全违背教育测量学原理。现在情况情况已有改观。

综上所述,中国在医学院校考试方面差距相当大,这些问题,是我们立项的依据,也是改革的目标。

第二节 我国高等医学教育考试改革(认知能力评价项目)的过程与成果

从 1991 年 11 月至 1992 年 11 月,我们进行了第一阶段,即改革的准备工作。包括美国专家技术援助,组建研究队伍,对全国医学考试基本情况调查,学习翻译国外有关文献,出版专著,购置设备,赴美学术访问等。通过这些工作,对于我国医学考试与美国的医学考试进行了系统的比较,找出了差距和问题。

在第一阶段工作的基础上,我们于 1992—1993 年进行了第二阶段的工作,即研究与实验。结合我们的具体情况,引进美国的考

试方法,成功地在中国医科大学校内进行了考试的研究与实验。

第三阶段的工作,1993年11月,经请示卫生部科教司支持,在第二阶段研究实验的基础上,以自愿为原则,邀请了全国17所高等医学院校参加协作,进行医学考试方面的推广与应用工作。这17所院校是:大连医科大学、上海第二医科大学、山东医科大学、中国协和医科大学、中国医科大学、石河子医学院、白求恩医科大学、北京医科大学、福建医学院、华西医科大学、首都医科大学、重庆医科大学、第一军医大学、浙江医科大学、湖北医科大学、湖南医科大学、上海医科大学。17所院校经过认真考虑后,同意参加由中国医科大学主持的CMB91-547项目第三阶段联合考试,并派出院校长和教务处长参加项目协作会。在这次会议上就中国医学考试现状与国际标准的差距,第三阶段协作的目标、任务、工作计划进行了充分讨论,取得共识,并决定对17所医学院校高年级学生进行实习前和实习后两次统一的临床医学综合考试。通过联考达到推广与应用的目的。全国联考的具体任务为:

1. 推广医学考试命题的科学化方法

命题是考试的重要环节,其质量高低直接关系着考试的成功与失败。因此,为使教师能出高质量的试题,中国医科大学按国家教委制定的“课程基本要求”制定了考试大纲,按照考试大纲设计了具体的命题方案。命题方案按学科、内容、题型、认知分级、大纲要求、预计难度等6个方面的要求来制定,然后各校命题委员会将自己学科试题按命题方案进一步具体到每一试题。即每一试题的学科归属、测试内容、认知分极程度、大纲要求的程度、预计难度等都要确定,这样确定的试题,即保证了试题质量,又避免了教师主观因素的影响,从而提高了考试的效度和信度。为以后建立大型试题库奠定了基础。

2. 医学考试组织管理科学化

全国范围考试的组织和实施是一个系统过程,每一个环节的失误都会给考试结果带来误差,因此,必须制定从命题到阅卷每一